DOCUMENT RESUME

ED 109 261 95 TM 004 730

AUTHOR	deAvila, Edward A.; Havassy, Barbara
TITLE	I.Q. Tests and Minority Children.
INSTITUTION	Dissemination Center for Bilingual Bicultural Education, Austin, Tex.; Multilingual Assessment Program, Stockton, Calif.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
PUB DATE	Jan 74
NOTE	40p.
AVAILABLE FROM	Dissemination Center for Bilingual Bicultural Education, 6504 Tracor Lane, Austin, Texas 78721 ($1.00)
EDRS PRICE	MF-$0.76 PLUS POSTAGE. HC Not Available from EDRS.
DESCRIPTORS	Achievement Tests; Computer Programs; Culture Free Tests; Elementary Education; Individual Tests; *Intelligence Quotient; *Intelligence Tests; *Minority Group Children; Models; Norms; Program Evaluation; Socialization; Student Evaluation; Test Bias; *Testing Problems; Test Interpretation
IDENTIFIERS	Elementary Secondary Education Act Title VII; ESEA Title VII; PAPI; Piagetian Tasks; *Program Assessment Pupil Instruction System

ABSTRACT

Test publishers and the users of standardized IQ and summary-score achievement tests have failed to consider the problems associated with testing the minority child. Since the results of these tests are used to determine the education, economic, and social future of children, a harder look at the minority child's cultural background is essential. The problem of testing cannot be solved by recreating standardized tests for minority children which are based on old conceptions of intelligence and educational achievement. A radical change in the whole approach to testing and the generation of entirely different models of education and of testing is needed. The Program Assessment Pupil Instruction (PAPI) System is one step in an attempt to move in a new direction. The PAPI uses four Piagetian-based measures: Cartoon Conservation Scales; Water Level Task; Figural Intersections Tests; and Serial Task. This system can be used with any child to produce educational program data, or data for program evaluation. Given the problems associated with testing, the PAPI system suggests one approach to the issue which speaks to the different needs of people within the educational community. (RC)

ED109261

TM

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Juan D. Solis

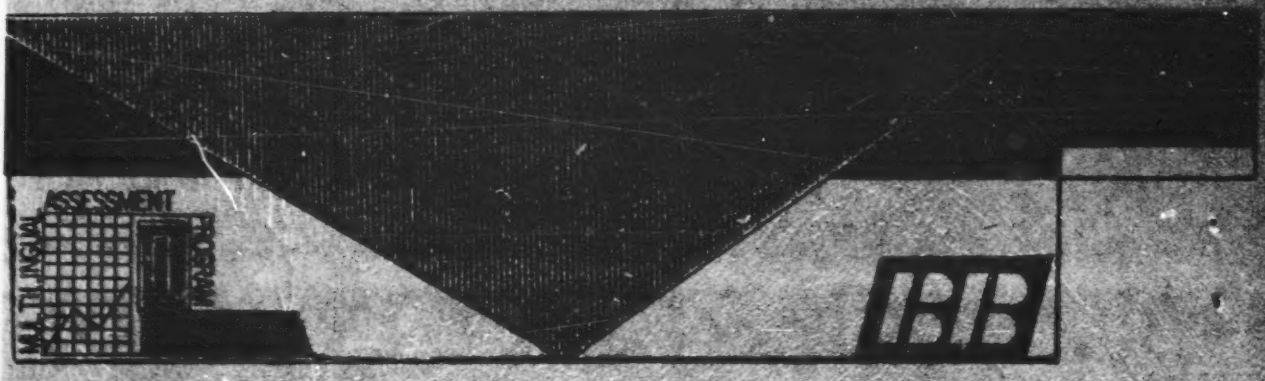
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

U.S. DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

I.Q. TESTS AND MINORITY CHILDREN

TM 004 730



I.Q. TESTS

AND

MINORITY CHILDREN

Edward A. de Avila

and

Barbara Havassy

January 1974

Developed by

Multilingual Assessment Program

Stockton, California

Published and disseminated

by

DISSEMINATION CENTER FOR
BILINGUAL BICULTURAL EDUCATION

6504 TRACOR LANE AUSTIN, TEXAS 78721

"The project reported herein was performed pursuant to a Grant from the U.S. Office of Education, Department of Health, Education and Welfare. However, the opinions expressed herein do not necessarily reflect the positions or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred."

The Dissemination Center for Bilingual Bicultural Education is a special Title VII ESEA project funded by the U.S. Office of Education through Education Service Center, Region XIII. The Center has selected these materials for dissemination however the opinions expressed herein do not necessarily reflect its position or policy nor that of Education Service Center, Region XIII.

This publication was developed and printed with funds provided by Title VII of the Elementary and Secondary Education Act of 1965. Therefore, it is in the public domain and may be reproduced for local use.

CONTENTS

FOREWORD. v

ACKNOWLEDGEMENTS. vii

Traditional Tests 1

Ethnic Norms. 3

Culture-free Tests. 4

The Validity and Utility of the I.Q. Score. 9

Informational Needs Within the Educational System and the I.Q. Tests 11

PAPI: A Piagetian-based Alternative. 19

SUMMARY 29

REFERENCES. 31

Dissemination Center
for
Bilingual Bicultural Education
6504 Tracor Lane
Austin, Texas, 78721

Juan D. Solís, Director

Education Service Center, Region XIII
Austin, Texas, 78721

Joe Parks, Executive Director

Royce King, Director, Division of
Program Development

FOREWORD

This publication, entitled *I.Q. Tests and Minority Children*, demonstrates point by point the inadequacies of IQ testing for Spanish-speaking children and children of other minorities.

Designed for use by educators at every level, this edition provides a good deal of useful information about tests of intelligence based on translations, ethnic norms, and other elements that are not equally familiar to minority group children.

The authors of this publication not only present the problems involved in IQ testing, they also provide suggestions for solving these problems; for example by implementing specific educational programs with computer-based models, such as the Program Assessment Pupil Instruction (PAPI) which uses Piagetian-based measures specifically designed for each child in the classroom.

This book and others have been developed by the National Multilingual Assessment Program, a specially funded project under P.L. 89-10, Elementary and Secondary Act, as amended, Title VII.

The Dissemination Center for Bilingual Bicultural Education is indebted to the Multilingual Assessment Program, Joe Ulibarri, Director, for allowing us to publish and disseminate the first edition of this book. We feel that this work will help to effect needed changes in the method by which achievement is measured in minority group children, especially Mexican-American students.

Requests for information concerning this book and other bilingual materials should be addressed to the Dissemination Center for Bilingual Bicultural Education, 6504 Tracor Lane, Austin, Texas 78721.

Juan D. Solís
Director

ACKNOWLEDGEMENTS

The authors wish to express their deep gratitude to Dr. Juan Pascual-Leone of York University, Toronto, Ontario, Canada, for his critical review of several drafts and for the use of his Water Level and Figural Intersection tests in the test battery. They would also like to acknowledge Veronica Hooker's work on the computerized classroom activities.

TRADITIONAL TESTS

Traditional tests of intelligence are inappropriate for minority children, particularly children of non-English speaking backgrounds. Such diverse groups as the National Education Association, popular press, courts, civil rights organizations, state and federal agencies and school psychologists have all pointed to the failure of the test publishing industry to fully consider the cultural and linguistic differences of minority children when constructing, publishing and selling tests.

Publishers have responded to this criticism by:

1) translating existing intelligence tests for non-English speaking children;
2) adjusting norms for ethnic sub-groups;
3) attempting to construct culture-free tests.

There are distinct problems with each of these approaches. In addition, there are problems concerning the basic validity and utility of information produced by IQ tests. A discussion of the inadequacy of the response of test publishers and a presentation of several other issues follows.

TRANSLATIONS

Translating existing intelligence tests for non-English speaking children often creates more problems

than it solves. Regional differences within a language make it almost impossible to use a single translation. Thus, while the word "tostón" refers to a half dollar for a Chicano child, for a Puerto Rican it refers to a squashed section of a fried banana. Mono-lingual translations are also inappropriate because the language familiar to non-English speaking children is often a combination of two languages as in the case of "pocho" or "tex-mex". Furthermore, many non-English speaking children have never learned to read in their spoken language. One finds many examples of tests written in Spanish being given to Chicano children who may speak Spanish but have had no prior instruction in reading Spanish.

Another problem in translating tests is that the direct translation of a word or phrase in one language may result in a word which is not used with the same frequency or have the same potency in the second language. For example, the word pet is a common word in English. Its Spanish equivalent, animal doméstico, is almost never used. Also, translating a word from one language to another can vastly alter its meaning, like the wide variety of seemingly harmless English words

which translate into Spanish swear words or "palabras verdes". Thus, translating a large egg into a "huevón" may satisfy grammatical requirements and seem harmless to a translator, but it has a more earthy connotation for Chicanos and Puerto Ricans.

ETHNIC NORMS

The second major response of the testing industry to criticism has been to establish or propose re-establishment of regional and ethnic norms; in other words, to "compensate" minority children for their "deprived backgrounds". Not only will such a practice lead to lower expectations for minorities (which, in turn, lowers children's aspirations to succeed), it is as shortsighted as awarding Chicano children extra points because "they speak a little Spanish." Ethnic norms take no account of the complex reasons why minority children on average score lower than Anglos on IQ, and they are potentially dangerous because they provide a basis for invidious comparisons between different racial groups. The tendency is to assume that lower scores are ultimately indicative of lower potential, thereby contributing to the self-fulfilling prophecy of lower expectations for minorities as well as reinforcing the genetic inferior-

ity argument advanced by Jensen (1971) and others.

Furthermore, if test publishers and users are willing to consider the establishment of ethnic norms, similar arguments could be made for the establishment of norms based on sex differences as well. Considering sex and ethnicity would require an almost infinite set of different norm tables in order to account for all of the different ethnic subgroups in the United States. From the practical point alone this leads to an absurdity. Finally, the establishment of ethnic norms assumes that the groups are ethnically homogeneous with little or no cross-over or intermarriage. One might wonder what publishers would propose to do with a set of male/female twins who had a Mexican father and a Hungarian mother?

CULTURE-FREE TESTS

Another way in which the testing industry has responded to criticism of conventional IQ tests has been to create"culture-free" tests. Such tests are difficult, if not impossible, to construct. In tests of mental ability an attempt is made to determine the ability of a child to manipulate certain elements of a problem into a predetermined solution. But if all or some of the

elements are not equally familiar to the child, the test is unfairly biased. The influence of culture on conventional IQ test items is subtle in some cases, blatant in others. But the fact remains that in a large number of traditional IQ tests, the items are measuring something other than that for which they were designed. Items particularly influenced by cultural factors fall into the following general classifications:

Socialization. Items of this type are couched in such a way as to actually be measures of the child's family value system. The referent system is, of course, the dominant Anglo-American middleclass. The confounding effects of this problem are particularly evident in the "Comprehension" scale of the Weschler Intelligence Scale for Children (WISC) where children were asked such questions as:

> "What is the thing to do if you lose one of your friend's balls? or "What is the thing to do if a fellow much smaller than yourself starts a fight?"

Allowing for the stilted manner in which the questions are phrased and assuming that the child knows all of the vocabulary, it still seems perfectly obvious that

this type of question has little or nothing to do with a child's ability to process, manipulate and/or code information; the answers depend almost exclusively on whether a child has been socialized under the particular ethical system implied by the question.

Productivity or Level of Aspiration. Many tests confound what they hope to measure with a measure of productivity or level of aspiration. For example, in a large number of tests the child who produces the largest number of responses is rewarded while the child who stops responding after only a few attempts is punished by receiving a lower score. Thus, in the Draw-A-Man test the child who produces the more elaborate figure stands the better chance of receiving the higher score, the assumption being that all subjects will produce as many responses as they are able--that they all have the same level of aspiration.

Timed tests and "endurance" tests also fall prey to a confusion between the measurement of ability and the measurement of aspiration. In timed tests, which constitute the majority of published group tests, the children are asked to work quickly, quietly and efficiently without regard for the child who is simply not in a

hurry and not particularly motivated to be so. The endurance test, for the purpose of boosting statistical reliability, requires that the child answer a large number of questions which vary little in content. This problem is particularly evident in group tests such as the Lorge-Thorndike and the California Test Bureau series. Similar to the "endurance" test is the test where items are sequenced in order of increasing difficulty. This design feature characterizes most of the standardized tests. In this situation the child is forced--by design--to encounter increasing levels of failure and frustration. In the case of the child who starts out fearfully, as do most minority children, the first indication of failure or difficulty is enough to discourage him or her from continuing and one finds the child "staring blankly off into space".

Experience or Specific Learning. In tests which require answers of fact, there is an implicit assumption that all the children taking the test will have had an even chance, more or less, of having been exposed to the facts being tested. The spuriousness of this assumption is witnessed by any number of examples where children are asked questions of vocabulary. In such instances,

it is impossible to determine whether a minority child has missed a test item because he lacks the capacity to understand a given word or because he simply has never been exposed to the word. Nitroglycerin (in the WISC), fire hydrant (in the Betty Caldwell Preschool Inventory and Peabody Picture Vocabulary Test) or crevice (in the Otis-Lennon) are terms unlikely to be heard or used in the home of the average low SES or minority child.

The WISC, perhaps the most widely-used individually-administered intelligence test in the world, is replete with examples of the importance of specific experience on test results. Take the WISC "General Information" item, "In what kind of store do we buy sugar?" If a child lives in an Indian reservation, he might buy his sugar in a drug store or at a trading post. A Chicano child might reply to that question, "at the Chinitos", a small variety shop owned by a Chinese family. Yet these are not acceptable responses for credit in the WISC. Consider another WISC item, "Where is Chile?" What if a child eats chile, as a Chicano child might. Or consider the item, "What is the thing to do if you lose one of your friend's balls?" The acceptable WISC responses are "Give him one of mine. . .try to get it

back or replace it. . .or try to find it." A probable reaction of a Barrio or inner city child (depending on the size of the other child) would be avoidance, to escape the child's anger. That is not to say that these kinds of experience-based items are only found in the WISC scale. They are examples of these kinds of biases found in almost all tests of intelligence.

THE VALIDITY AND UTILITY OF THE IQ SCORE

The basic justification behind the use of the IQ score is that it statistically predicts to mental retardation and low achievement. In fact, the IQ test is the sine quo non for screening children suspected of mental retardation. Mercer (1971) found that of those persons who would have been labelled as mentally retarded if their classification depended solely on test scores, a full 84% had completed 8 grades or more in school, 83% had held a job, 80% were financially independent or a housewife, and almost 100% were able to do their own shopping and to travel alone. In other words, it is probable that even at the task for which experts agree the IQ test is best suited--screening for mental retardation--the IQ measure has a dubious real life validity.

The IQ test is also considered by many educators

and politicians to be a useful instrument for teachers--for discovering unnoticed learning problems or intellectual strengths, for example. Indeed, many states mandate that districts administer IQ tests several times in a child's succession through the school system. But does the result really help the teacher? Let us take a typical example. A teacher suspects a child of having a severe learning disability. She asks the school or district psychologist to test the child. The psychologist gives the child the WISC in which he scores, let us say, an IQ of 87. This psychologist happens to be extremely conscientious, so he devotes the next few days to writing up an extensive report of his impressions of the child's performance and potential. The child comes from a poor background, the psychologist writes. He has many sibs and is low in the hierarchy with regard to getting attention. After much reviewing the child's school file and his own notes, the psychologist writes a report and hands it to the teacher. The teacher responds in surprise, "But I knew all that. What I want to know is what to do; how can I teach this child." Thus, in most cases neither the psychologist nor the teacher are any wiser despite considerable time and expense administering and

evaluating the IQ test. In this situation the psychologist often finds him or herself in the middle of the delicate quasi political-social balance between the classroom teacher, the principal, the district administration, the child and his family.

While few psychologists would agree with the notion that educational decisions affecting the life of the child should be made exclusively on the basis of a single full scale IQ score, the fact nevertheless remains that these decisions are made by educators who, through personal fiat supported by state mandate, ignore both the individual sub-scale profiles as well as psychologists' admonitions, simply for the sake of practical expediency. The results is, of course, a form of default institutional racism.

INFORMATIONAL NEEDS WITHIN THE EDUCATIONAL SYSTEM AND THE IQ TESTS

Much of the controversy surrounding IQ tests and minority children focuses on whether the IQ model is a valid one. A more practical, and less abstract issue which needs to be answered concerns the general utility of the information produced by the test, i.e., what can one do with the information provided by such a test? In

order to answer this question we must consider who is asking the question and for what purpose. It should be apparent that there are qualitative differences concerning the type of information needed within the educational system. These differences depend on the source of the need. To a large extent, much of the confusion surrounding the issue of whether to test stems from a failure to consider these differences. A consideration of the qualitative difference between these different needs will hopefully lend some clarity to the general controversy as well as serve to introduce a procedure that attempts to meet some of the specific needs of each.

There are three general levels of organization within the educational system that require information traditionally obtained through IQ testing. These three may be described as (1) the funding level which involves educational agencies supra-ordinate to the local school districts; (2) the local level which consists of both district personnel and building-principals and finally; (3) the classroom teacher, paraprofessional and parent. Whether or not the IQ score information can serve any ultimate useful function will depend on the particular and peculiar needs of each of these three groups. It

is unfortunate that historically all three attempt to use the IQ score for radically different and often conflicting purposes.

Consider first the supra-ordinate funding agencies such as state and federal departments of education. For purposes of determining the allocation of funds, these agencies require information concerning educational and program needs at the school district level. While one source of information they generally use are IQ measures, one wonders whether the information produced by IQ tests adds anything of value to needs assessment. More appropriate "needs assessment" procedures, it would seem, would be restricted to the assessment of whether specific educational programs in areas of reading, arithmetic or the like are needed rather than attempting to infer specific need from omnibus assessments based on so poorly understood a concept as IQ.

The supra-ordinate agencies have a second related informational need. In contrast to needs assessments which very often can be conducted through examinations of school attendance records, age-grade placement patterns and achievement data in the broad sense, funding agencies need to know about the effectiveness of particular programs.

Due to "accountability" and "evaluation/audit" requirements, the agencies chiefly responsible for the allocation of funds have mandated that testing be conducted at the child level as a means of collecting information which can be used to evaluate educational programs. In actuality, program evaluations can be made through a variety of means or procedures, none of which necessarily has anything to do with IQ or summary-score types of measures discussed above. For example, a reasonable assessment of program effectiveness can be made through the collection of nominal data at various levels of school and community organizations. Thus, administrators, teachers, parents and children can be interviewed as to their perceptions of the effectiveness of the different processes of the program. On the other hand, more product-oriented assessments, at the child level, can be made according to the specific educational objective of the program. The reporting of these latter data can be made on the basis of group change scores, without reference to individual scores, eliminating the potential dangers inherent in individual scores. Finally, all of the various data can be integrated to produce a report which describes the program and weights the relative importance of each

aspect of the program in recommending changes in various parts of the total operation.

It is interesting to note that there is a paradoxical element in using IQ scores to evaluate program-related change. The paradox stems from the fact that the IQ model is based on the notion that intelligence is static and hence not subject to change. What this means is that educators have been using IQ scores to evaluate change which can not occur according to the IQ model. The IQ model, which negates time or age through the division of mental age by chronological age, is by definition a static model and, therefore, inappropriate for measuring program-related change which must take place over time.

At the local level, school district personnel need information as to the needs of children and the effectiveness of individual programs in the same way as do the supra-ordinate agencies. However, needs assessments are usually conducted at the state level and school officers more often than not use the state-provided information as a statement of need, preferring not to conduct detailed expensive in-house initiated research.

Ideally, the evaluation of individual programs should center around the collection of data which follows directly

from the particular program objectives and activities. However, it is often the case that instruments of evaluation have very little to do with actual programs. Rather than program-specific measures, very often IQ or nationally normed achievement tests are used. Since IQ and other such scores provide precious little in the way of information about the effectiveness of individual programs and program components, there results great confusion in interpreting program evaluations. For example, in comparison with an appropriate "control group" suppose that children who participated in a special program on arithmetic computation had received higher scores on a group IQ score than they had the year before. With this type of evaluation officials would be hard put to say exactly which aspects of the program had led to the improved IQ scores since the connection between the two remains either ill-defined or nonexistent. However, in contrast, a test (administered before and after a specific program) which sampled from each of the different aspects of the program would provide specific information on which aspects of the program led to improved scores and which did not. In other words using IQ scores to evaluate programs not specifically designed for the purpose of elevating IQ

scores is a little like using a thermometer to record height changes in a given period of time.

Unfortunately, the last to be considered in the educational hierarchy is the classroom teacher and what she needs in order to assist the learner. What can the teacher do with an IQ score? The limited utility of the IQ or for that matter any other summary score whether it be reading level according to grade equivalent score, grade point average, percentile rank or the like has already been discussed. Perhaps the fact that to a large extent the tests do not provide the teacher with much usable information is historically related to the fact that psychological tests were initially developed for purposes not functionally related to the present daily needs of the classroom teacher. It is also of some interest to note that even to the present day it is not the classroom teachers who either design the format of the information produced by the testing or, select the specific test or tests which are to be used.

There is still another way in which the use of IQ scores in the classroom might be questioned. Given it is possible to avoid the problems associated with adopting tests which employ the IQ model there is a final question

at the teacher-level concerning the ultimate functional use of the IQ score. If it is given that the IQ score is an accurate index of the children in a classroom, then what? What a teacher is asked to do is to discriminate between scores in designing curriculum for each and every child on the basis of information which she can not discriminate. What would a teacher do differently for a child with an IQ of 92 that she would or would not do for a child with an IQ of 100? One might agree that these two scores are functionally equivalent because they are both within the normal range (i.e., within one standard deviation of the mean). However, what about the case of a child who has an IQ score of 84 which is approximately a standard deviation below the mean and in some states considered as possibly falling into "EMR", "retarded" or "slow learner" categories? In the one case where the comparison is made between 92 and 100 there would be no difference in educational recommendation, however, in the other case where scores of 92 and 84 are compared, very different educational recommendations would be made. In both cases there is the same absolute difference in scores.

The point is that we have been basing educational

decisions on IQ tests which often produce questionable results. In many cases, the same criticisms apply to achievement tests which provide collapsed scores. Functionally, what programmatic difference can the teacher infer from scores provided in grade equivalent terms. What are the different educational decisions to be drawn from reading grade equivalency scores of 3.2 versus 3.6 and 3.6 versus 4.0? In the case of both the IQ score and the collapsed achievement score, the teacher, who is ultimately the person held "accountable", is simply not being provided with enough information upon which to base sound daily educational decisions. The teacher is thus by default forced to make arbitrary decisions which in most cases can not be defended. The fact that the tests have "predictive validity" in that they forecast failure on the part of Spanish-speaking or other minorities and success for the majority only speaks to the inappropriateness of the curriculum and to the failure of the American educational system to provide education for many of its children.

PAPI: A PIAGETIAN-BASED ALTERNATIVE

In the following, a computer-based model will be described which was designed with many of the above con-

cerns in mind. This model has been successfully tested in a variety of ways in three Southwestern and two Western states. The discussion provided below will outline the elements of the model. More substantive discussion concerning the psychometric features of the measures the model employs are subjects of several other articles currently in the journals or being prepared for publication.

The program Assessment Pupil Instruction (PAPI) system uses a number of Piagetian-based measures to generate two basic types of information by means of a centralized computer data-processing program.

The first type of information provided by the PAPI system is statistical in nature and is meant for program evaluation at the funding, administrative and district levels. The second type of information is oriented towards the needs of the teacher and consists of recommended activities specifically designed for each child in the classroom.

In its present form, PAPI uses four Piagetian-based measures developed jointly and individually by Pascual-Leone and DeAvila. Two types of measures are used. The first is used as a base measure of achievement where information is collected which follows from classroom

activity. The second type of measure attempts to provide an index of the child's current level of development and involves three different tests.

In the following section, each of the four tests in the battery is described. The first to be described is the measure employed as the basic achievement index. The following three are used to provide an index of level of development. The description of the tests is followed by a more detailed description of the PAPI system.

1. CARTOON CONSERVATION SCALES (CCS)

Several measures of Piaget's conservation tasks are assessed by means of a cartoon format developed by De Avila, Struthers and Randall (1969). In De Avila's procedure, three horizontal cartoons are presented in which two children are discussing the task. In the first frame an equality is established between two objects according to the dimension being tapped (i.e., number, length, substance, etc.). In the second frame an identity transformation is depicted and in the third frame the question of equality is asked. On the right side of the panel three possible answers are presented in a vertical order. The three alternatives which show the characters responding to the question are randomly ordered as to correctness in order to avoid position effects. Similarly, within a scale wording is altered from panel to panel in order to avoid the possible effects of acquiescence. Background on the conservation scales and an illustration of the dialogue from each scale are presented below.

Struthers and De Avila (1967) and De Avila et al (1969) and others have validated the CCS procedure in a variety of ways. In one study De Avila, et

al (1969) tested thirty male and female first grade subjects using both group and individual clinical methods. Statistically significant correlations were obtained across the two methods of assessment ($\underline{r}$ = 0.663). The scales also possessed a high degree of internal consistency as shown by an examination of factor analytic structure and reliability indices.

In its current form the CCS consists of thirty cartoon panels; six examples of five tasks. The panels are presented to the subjects and the story line is read and elaborated upon in order to facilitate understanding of the question. The subjects task is to mark the one (alternative) "that makes the story true."

2. WATER LEVEL TASK (WLT)

The conservation of the horizontality of water described by Piaget and Inhelder (1948), and used by Smedslund (1963), Dodwell (1963), Rebelsky (1964), Beilin, Kagan and Rabinowitz (1966) and Pascual-Leone (in press). This task involves the subject being able to break perceptual set to recognize that no matter what angle a bottle placed on a flat horizontal surface is viewed from, the water level will always be parallel to the horizontal surface. A more complete description of the relative parameters of this type of task can be found in the semantic-pragmatic analysis of Pascual-Leone (in press).

A special version of Pascual-Leone's group tests developed by Pascual-Leone and De Avila is used. Subjects are presented with individual booklets which contain five horizontal or vertical two-dimensional bottles, eight two-dimensional-tilted bottles and four three-dimensional bottles, two of which are also tilted. The subject is asked to draw a line where the top of the water would be if the bottle were half full and then to place an "X" in the part that contained the water.

3. FIGURAL INTERSECTIONS TEST (FIT)

The figural intersection test is a group administered paper-and-pencil test in which subjects are required to place a dot in the intersection space of a varying number of geometrical figures. It was developed by Pascual-Leone and constitutes a figural analogue to Piaget's work on intersection of classes. In a series of unpublished studies, Pascual-Leone has shown the test to have a high degree of internal consistency (split-half reliability = .89) as well as being significantly related to tests of similar logical structure (Pascual-Leone and Smith, 1968). For example, it has shown a high correlation with the WLT described above. Combined with the WLT, and ST described below, in the present context, it is taken as an index of developmental level. This relationship has been previously found in a series of unpublished studies by Pascual-Leone and Parkinson at York University, and De Avila (1971).

4. SERIAL TASK (ST)

The serial task (De Avila, 1971) is a short term memory task which is individually administered in two phases. First, subjects are pre-exposed to the stimulus materials used in a second testing phase. In the pre-exposure or pre-training phase, each subject is shown a series of 10 different 35mm color slide transparencies of pictures depicting a donkey, house, airplane, etc. Subjects sit facing a screen situated on a wall six feet away. The 10 illustrations are presented by means of a Kodak 650 carousel slide projector. To introduce the task, each subject is shown each figure and asked to give its name and color (i.e., "a yellow hat"). Following this initial introductory phase and after the subject is able to correctly identify each figure ten times when presented in rapid random succession, the testing phase takes place.

The test phase is conducted in a "free recall" manner where, without any prior knowledge of the length of a list, the subject is asked to reproduce the list ignoring the order in which the individual items

are presented. Subjects are shown a series of individually presented figures terminated by a blank slide, and asked to tell the experimenter what they have seen. The exposure time for each individual slide is .750 msec. There is no requirement that the sequence of the presentation be maintained, or that the subject respond within a specified period of time, or produce a predetermined number of responses. The child is simply asked to reproduce what he has seen using whatever labels are convenient.

There are seven sets of figures presented to each subject. These seven sets vary as to the number of stimuli within a series. The number of figures presented within a series, as well as the individual figures, are randomly varied. Finally, each illustration is presented no more than once within a series.

The PAPI system uses these four measures to produce several types of information each of which has a particular function. First, psychometric information on each test is produced as a means of assuring that the tests used in any given case have satisfactory levels of reliability and validity. This information is of primary importance to the researcher or program evaluator as a check against various types of problems associated with test administration, sampling procedures and so on. The second type of information produced by the PAPI system can be described as developmental-normative. These data provide normative information regarding the current developmental levels of the entire sample. Moreover,

these data can also be stored to serve as base-line data which can be compared across temporal (as in a pre-post design) or situational/program dimensions. In summary, these two types of information, the psychometric and the normative/developmental data, are geared to the needs of program evaluation and needs assessment at the administrative levels.

The third type of information generated by the PAPI system consists of lists of recommended classroom activities for each child tested. These activities are geared to providing children with experience in specific educational activities thought to be important in the overall educational/program objectives.

In generating the suggested classroom activities for each child, four factors are taken into account:

1) The child's achievement level with respect to the concepts measured by the CCS.
2) The child's developmental level as measured by the WLT, FIT and ST tests.
3) The achievement level of the child's referent group with respect to the concepts measured by the CCS.
4) The developmental level of the child's referent

group as measured by the WLT, FIT and ST tests.

An example of one computer printout which lists specific classroom activities is provided below. Using the four factors listed above, the following example would indicate that:

1. The child's overall developmental test profile is similar to children of the referent group who have already acquired the concept.
2. The child does not understand the concept.

It is then inferred that a specific set of activities are developmentally/educationally appropriate with respect to the concept of number as tested by the CCS. For example, this conjunction of conditions would lead to the following set of recommended classroom activities:

CONCEPT OF NUMBER

THIS CHILD DOES NOT HAVE A CONCEPT OF NUMBER, HOWEVER, HE IS READY TO LEARN. TRAIN HIM BY GOING THROUGH THIS SEQUENCE, USING CONCRETE ACTUAL OBJECTS ONLY.

A. SORTING AND CLASSIFYING OBJECTS BY
   1. COLOR
   2. TEXTURE
   3. SIZE
   4. WEIGHT
   5. NUMBER
   6. WHATEVER OTHER ATTRIBUTES YOU CAN THINK OF

B. COUNTING ACTIVITIES, SUCH THAT THE CHILD LEARNS...

1. SEQUENCE, NUMBERS INCREASE BY ONE EACH TIME

2. INVARIANCE

A. PATTERN, 5 OBJECTS BUNCHED AND 5 OBJECTS IN A ROW ARE BOTH 5

B. OBJECTS, 5 ELEPHANTS AND 5 FLIES ARE BOTH 5

3. SYMBOLS, LEARN THE ARBITRARY NAMES OF THE NUMERALS

C. CONCEPT AND LANGUAGE OF EQUALITIES AND INEQUALITIES

1. > . WHICH HAS MORE, IS BIGGER, IS LONGER, ETC.

2. < . WHICH HAS LESS, IS SMALLER, IS SHORTER, ETC.

3. = . WHICH HAS JUST AS MUCH, IS JUST AS BIG, IS JUST AS LONG, ETC.

The above is an example of what is sent directly to the teacher. It can also be sent to the home so that the parents are made fully aware of what the teacher is trying to accomplish with the child. In this way, possibly with some guidance from the teacher, the parent can participate in the child's education while supporting the teacher.

It should be noted that the PAPI system is designed

so that a child's peer or referent group can be designed according to any number of nominal descriptors such as grade, sex or ethnic group. Thus far, only chronological age has been used because of the importance of the age variable in Piaget's theory. It should also be noted that, in the present description of PAPI system, the CCS is treated as an achievement test rather than as a test of development as it was originally designed. Actually, any test of achievement can be used within the system as long as it is linguistically and culturally appropriate and is constructed in such a way that it can be used to generate suggested activities.

In summary, the PAPI system can be used with any child to produce educational program data, or data for program evaluation. The purpose of the present discussion has not been to suggest that the system is in any way a complete educational package. Further development is needed in creating test procedures and in training educators, on how to use this type of system. But given the problems associated with testing, the PAPI system suggests one approach to the issue which speaks to the different needs of people within the educational community.

SUMMARY

The major thesis has been that test publishers and the users of standardized IQ and summary-score achievement tests have failed to consider the problems associated with testing the minority child. And since the results of these tests are used to determine the educational and, by extension, the economic and social future of school-age children, it behooves test publishers and the educational community to take a harder look at the minority child's cultural background. It is the authors' opinion that consideration of these issues leads to the conclusion that the problem of testing cannot be solved by attempts to recreate standardized tests for minority children which are based on old conceptions of intelligence and educational achievement. It is concluded that what is required is a radical change in the whole approach to testing and the generation of entirely different models of education and of testing. There is growing support for this conclusion (for example see McClelland, 1973). The PAPI system is just one step in an attempt to move in a new direction.

REFERENCES

Beilin, H., Kagan, J. and Rabinowitz, R. Effects of verbal and perceptual training on water level representation. Child Development, 1966, 37, 317-328.

De Avila, E. A. The use of artificial language in the study of processing capacity, immediate memory and Piaget's cognitive development variable. Unpublished doctoral research, York University, 1971.

De Avila, E. A., Struthers, J. A. and Randall, D. L. A group measure of the Piagetian concepts of conservation and egocentricity. Canadian Journal of Behavioural Science, 1969, 1(4), 263-272.

Dodwell, P. C. Children's understanding of number and related concepts. Canadian Journal of Psychology, 1960, 14, 191-205.

Jensen, A. R. Do schools cheat minority children? Educational Research, 1971, 14, 3-28.

McClelland, D. Testing for competence rather than for "intelligence". American Psychologist, 1973, 28, 1-14.

Mercer, J. The meaning of mental retardation. In Richard Koch and James Dobson (Eds.), The mentally retarded child and his family: A multidisciplinary handbook. New York: Brunner/Mazel, Inc., 1971.

Pascual-Leone, J. and Smith, J. The encoding and decoding of symbols of children. A new experimental paradigm and a neo-Piagetian model. Journal of Experimental Child Psychology, 1968, 8, 328-355.

Piaget, J. and Inhelder, B. La representation de l'espace chez l'enfant. Paris: P.U.F., 1948. English translation: The Child's Conception of Space. London: Routledge and Kegan Paul, 1956.

Rebelsky, F. Adult perception of the horizontal. Perceptual and Motor Skills, 1964, 19, 371-374.

Smedslund, J. The effect of observation on children's representation of the spatial orientation of a water surface. Journal of Genetic Psychology, 1963, 102, 195-201.

Struthers, J. and De Avila. E. A. Development of a group measure to assess the extent of pre-logical and pre-causal thinking in primary school age children. Paper presented to National Science Teacher's Association Convention, Detroit, 1967.

STAFF

Juan D. Solís, Director
Joanna F. Chambers, Research Librarian
Elsa Sanchez de la Vega, Editor
Blanche Armendáriz, Editor
Pam Harper, Editor
Sarah Frey, Assistant Editor
Stephan Jackson, Evaluator
Suzanna Cortez, Dissemination Clerk Technician
Jan Anderson, Dissemination Clerk
Martha Basden, Secretary
Joanna Melcher, Typist
Judy Brown, Typist
Delia Hernández, Typist
Dahlia López, Typist